

Navigation conceptuelle dans une base de connaissances sur l'usage des plantes en santé animale et végétale

Priscilla Keip, Amirouche Ouzerdine, Marianne Huchard, Pierre Silvie, Pierre
Martin

► To cite this version:

Priscilla Keip, Amirouche Ouzerdine, Marianne Huchard, Pierre Silvie, Pierre Martin. Navigation conceptuelle dans une base de connaissances sur l'usage des plantes en santé animale et végétale. CORIA: Conférence en Recherche d'Information et Applications, May 2019, Lyon, France. lirmm-02091989

HAL Id: lirmm-02091989

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-02091989>

Submitted on 6 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Navigation conceptuelle dans une base de connaissances sur l'usage des plantes en santé animale et végétale

**Priscilla Keip¹, Amirouche Ouzerdine², Marianne Huchard²,
Pierre Silvie^{1,3}, Pierre Martin¹**

1. CIRAD, UPR AIDA, F-34398 Montpellier, France

AIDA, Univ Montpellier, CIRAD, Montpellier, France

priscilla.keip@cirad.fr, pierre.silvie@cirad.fr, pierre.martin@cirad.fr

2. LIRMM, Université de Montpellier, CNRS, Montpellier, France

marianne.huchard@lirmm.fr

3. IRD, UMR EGCE, F-91198 Gif-sur-Yvette, France

RÉSUMÉ. De plus en plus de bases de connaissances sont développées dans le domaine de l'environnement et du vivant. L'enjeu de ce travail est de rendre ces connaissances accessibles à des utilisateurs ayant des préoccupations variées. Notre cas d'étude concerne une base de connaissances rassemblant environ 30 000 descriptions d'usage de plantes en Afrique, à effet pesticide, antimicrobien et antiparasitaire de synthèse. Dans cet article, nous proposons une architecture de système pour explorer et naviguer dans les connaissances grâce à l'analyse de concepts formels et proposons un nouvel algorithme faisant émerger un nouveau contexte d'emploi d'une plante dans un cas d'utilisation.

ABSTRACT. More and more knowledge bases are being developed in the field of environment and life. The challenge of this work is to make knowledge accessible to users with various concerns. Our case study deals with a knowledge base containing 30 000 descriptions of plant use in Africa, that are alternatives to synthetic pesticide and antimicrobial. In this paper, we propose a system architecture to explore and navigate within knowledge using formal concept analysis and present a new algorithm that reveals a new plant use in a context different from the initial use.

MOTS-CLÉS : classification, ontologie, exploration, navigation, base de connaissances, analyse de concepts formels

KEYWORDS: classification, ontology, exploration, navigation, knowledge base, formal concept analysis

1. Introduction

De plus en plus de systèmes à base de connaissances¹ sont développés dans le domaine de l'environnement et du vivant. L'enjeu est de proposer des méthodes et outils permettant de rendre ces connaissances accessibles à des utilisateurs ayant des préoccupations variées.

Notre cas d'étude est la base de connaissances Knomana issue du projet KNO-MANA (KNOWledge MANAgement on pesticidal plants in Africa for a safer food and a better environmental health)² et constituée récemment à partir de la littérature, dans le but d'identifier des solutions alternatives aux pesticides, antimicrobiens et antiparasitaires de synthèse et aux antibiotiques. Cette base est structurée par une ontologie et rassemble environ 30 000 descriptions d'usages de plantes en Afrique. Chaque usage, correspondant à une connaissance dans la base Knomana, est décrit au moyen de 36 informations. Parmi ces informations se trouvent la plante utilisée pour la protection, l'organisme cible (insecte, pathogène, etc.) et le système protégé (i.e. culture agricole, animal ou être humain). La base Knomana a pour vocation d'aider différentes catégories d'utilisateurs (agriculteurs, pisciculteurs, chercheurs, consultants, etc.) à répondre à des questions relatives à la santé animale et végétale. Ces questions sont diverses et on pourra trouver autant de questions que de besoins de la part des utilisateurs. Selon le cas, ces questions ou besoins demandent d'extraire des connaissances déjà présentes dans Knomana, à l'exemple de la liste de plantes permettant de lutter contre un insecte / une maladie sur un territoire, ou la liste des différents usages d'une plante, mais elles peuvent également nécessiter des analyses plus approfondies.

L'objectif du travail est d'élaborer un système de navigation et d'exploration dans les connaissances contenues dans Knomana. La navigation a pour objet d'identifier une connaissance présente dans la base à partir d'un ensemble de critères, et l'exploration d'établir une connaissance via une analyse qui satisfasse un ensemble de critères (Munzner, 2014). Pour organiser les connaissances, la méthode adoptée est une extension de l'analyse de concepts formels (ACF) (Ganter, Wille, 1999). L'ACF permet de grouper des objets en fonction de leurs attributs communs, puis d'organiser les différents groupes obtenus (appelés concepts) dans une structure hiérarchique (appelée treillis). Avec nos données, l'ACF peut, par exemple, classer des plantes de protection ou des systèmes à protéger en fonction de leur taxonomie et/ou de leur territoire. L'existence de relations entre types d'objets au sein des données a conduit à utiliser une version étendue de l'ACF : l'analyse de concepts relationnels (ACR) (Hacene *et al.*, 2013). La question de la visualisation des connaissances est traitée dans un autre cadre.

L'ACR étend les treillis produits par l'ACF en utilisant les relations binaires présentes entre les types d'objets. L'interconnexion entre deux concepts de treillis diffé-

1. Pour plus d'information sur la différence entre donnée, information et connaissance, se référer à l'article (Zeleny, 2013).

2. <https://ur-aida.cirad.fr/nos-recherches/projets-et-expertises/knomana>

rents matérialise les liens entre les objets contenus dans les concepts. La version de l'ACR utilisée dans ce travail est celle implémentée dans le logiciel RCAExplore³ qui permet de construire et de visualiser sous forme de graphes un ensemble de treillis interconnectés par des relations binaires. Son adoption pose deux problèmes. Le premier concerne la modélisation de la relation ternaire centrale de la base de connaissances au moyen de relations binaires. Le second concerne la visualisation au moyen d'un idiome qui n'est pas adapté pour les grands jeux de données.

L'objectif de cet article est de poser la problématique soulevée pour répondre aux objectifs pluriels de navigation et d'exploration de la base de connaissances structurée par une ontologie au moyen de l'ACR. La section 2 présente l'architecture du système qui permettra d'explorer et de naviguer dans les connaissances. Cette architecture met en évidence les constituants pré-existants. La section 3 présente un cas concret d'utilisation, celui demandant de faire émerger, à partir des connaissances contenues dans la base, une solution d'emploi d'une plante dans un contexte différent de l'usage initial. Un nouvel algorithme permettant de répondre à ce cas d'utilisation est proposé. La section 4 introduit les travaux existants dans les différents domaines abordés par ce travail et les pistes de recherche. Enfin, la section 5 présente les perspectives de travail.

2. Système de navigation conceptuelle

Le système logiciel comporte deux composants principaux inter-reliés, i.e. la base Knomana et un composant que nous appelons le *moissonneur conceptuel*.

Dans la base Knomana, chaque usage, correspondant à une connaissance, est décrit au moyen de 36 informations telles que le nom latin de la plante utilisée pour la protection, celui de l'organisme cible (insecte, pathogène, etc.), le système protégé (i.e. culture agricole, animal ou être humain), la méthode de préparation et d'application des extraits de la plante, l'effet sur l'organisme cible (par exemple toxicité ou répulsion), ainsi que la référence bibliographique d'où provient la connaissance. La particularité de la représentation des connaissances adoptée est qu'elle repose sur une relation ternaire connectant les trois informations principales entre elles : la plante de protection, l'organisme cible et le système protégé. Les questions relatives à cette base sont variées et possèdent différents degrés de complexité. Elles peuvent correspondre à des extractions de connaissances déjà présentes dans Knomana, ou nécessiter des analyses telles que (i) identifier des relations de protection qui existeraient entre un genre ou une famille de plantes et des genres/familles/règnes d'organismes cibles, (ii) connaître les plantes utiles pour protéger la plante mobilisée pour la protection d'une culture agricole, ou (iii) rechercher une solution d'emploi d'une plante dans un contexte différent de l'usage initial.

3. <http://dataqual.engees.unistra.fr/logiciels/rcaExplore>

La base Knomana est également constituée d'une ontologie et de dictionnaires de données (comportant notamment la classification du vivant) en plus du jeu de connaissances. Les dictionnaires de données et le jeu de connaissances sont actuellement implémentés dans un tableur et l'ontologie dans Protégé⁴. La FIGURE 1 présente la modélisation de la relation ternaire entre la plante de protection (appelée plante pesticide), l'organisme cible (appelé bioagresseur) et le système protégé sous forme de plusieurs relations binaires. Cette relation ternaire est matérialisée par une table principale (table *Connaissance*) reliée aux trois autres tables (à droite) au moyen des relations *ChasP* (*Connaissance has Pesticide*), *ChasB* (*Connaissance has Bioagresseur*) et *ChasS* (*Connaissance has Système Protégé*).

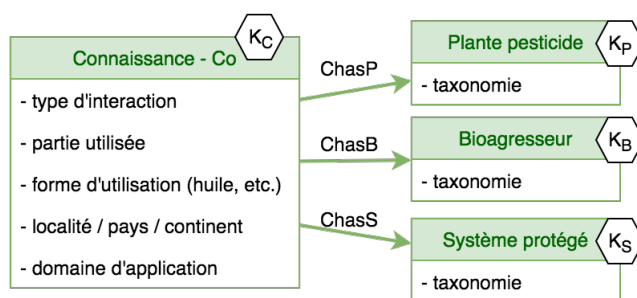


Figure 1. Représentation de la relation ternaire de l'ontologie, d'après (Keip, 2018).

Le moissonneur conceptuel, appelé *Conceptual Harvester* en FIGURE 2, comportera quatre composants :

- Le *Query builder* permet à l'utilisateur de formuler ses requêtes.
- Le *Knowledge extractor* assure la connexion avec Knomana et récupère les connaissances obtenues en réponse à la requête fournie par le *Query builder*.
- Le *Reasoning Tool* permet d'analyser les données. Ce composant comporte le logiciel RCAExplore (stéréotype UML Reasoning Tool). En partant de la requête et des données fournies par le *Knowledge extractor*, RCAExplore⁵ calcule les structures conceptuelles (treillis ou structures dérivées) en totalité, par étapes (Dolques *et al.*, 2013), ou à la volée à partir d'un groupe d'attributs ou d'objets d'intérêt (Bazin *et al.*, 2018). Par exemple, la FIGURE 3 présente une partie de chacun des treillis construits à partir du modèle de la FIGURE 1. Parmi ceux-ci, *Concept_Co_16* du treillis *Connaissances* rassemble les connaissances concernant (1) des plantes à effet pesticide (*ChasP*) listées dans *Concept_Ppest_19*, (2) des organismes cibles (*ChasB*) listés dans *Concept_BioAg_6* et *Concept_BioAg_2*, (3) des systèmes protégés (*ChasS*) listés dans *Concept_Pprot_4* et *Concept_Pprot_1* et (4) des connaissances du Burkina Faso (*paysBurkinaFaso*). RCAExplore devra être adapté à cette architecture logicielle car, outre le calcul, il propose actuellement une interface utilisateur et une visualisation non interactive des treillis.

4. <https://protege.stanford.edu/>

5. <http://dataqual.engees.unistra.fr/logiciels/rcaExplore>

– Le Knowledge viewer présente à l'utilisateur les résultats calculés par le Reasoning Tool sous forme graphique et permet l'interaction avec ce dernier pour la navigation ou l'exploration des connaissances.

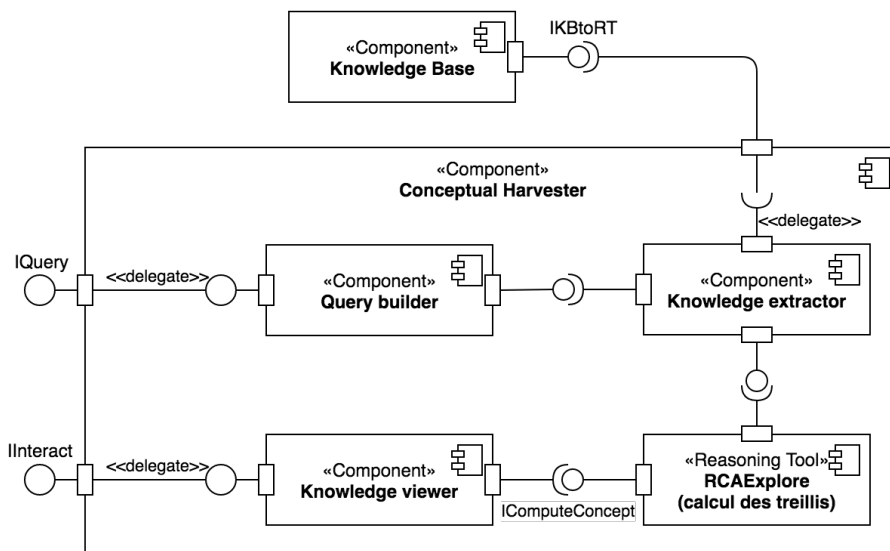


Figure 2. Diagramme de composants du système logiciel en utilisant le langage UML.

En attendant le développement du système logiciel, les éléments disponibles sont la base Knomana, RCAExplore et l'extracteur (appelé TableurToRcft) qui permet l'extraction des connaissances de la base et leur mise en forme pour alimenter RCAExplore.

La FIGURE 4 présente le diagramme de cas d'utilisation du composant Conceptual Harvester. Après avoir saisi une requête (Build query), potentiellement non restrictive, sur les connaissances, l'utilisateur navigue dans les connaissances (interact) via le Knowledge viewer selon différentes formes (navigate, launch reasoning tool, etc.). Il peut également modifier la requête (modify query) pour explorer les connaissances. Ces cas d'utilisation correspondent à la mise en œuvre des fonctions *explore* et *browse* de (Munzner, 2014). Ces cas étant génériques, leur spécialisation pour chaque type d'utilisateur (fermiers, sociétés privées, scientifiques, etc.) est envisagée.

3. Exemple de scénario d'utilisation

Dans cette section, nous nous intéressons à un scénario nécessitant l'exploration de Knomana. Pour illustrer ce scénario, la question considérée (Q1) est la suivante : *D'après Knomana, quelle plante pesticide pourrait lutter contre le Fusarium oxysporum au Burkina Faso ?*

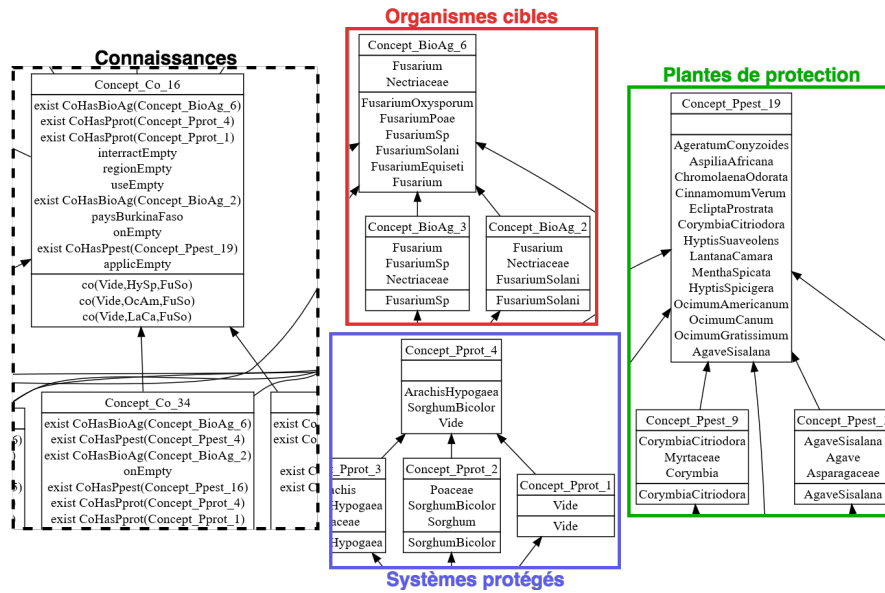


Figure 3. Parties des treillis concernant les données sur les *Fusariums*. Sur cette figure, un bloc représente un concept avec, de haut en bas, son nom, les propriétés partagées et les objets. Une flèche indique la relation d'ordre allant des sous-concepts vers les super-concepts.

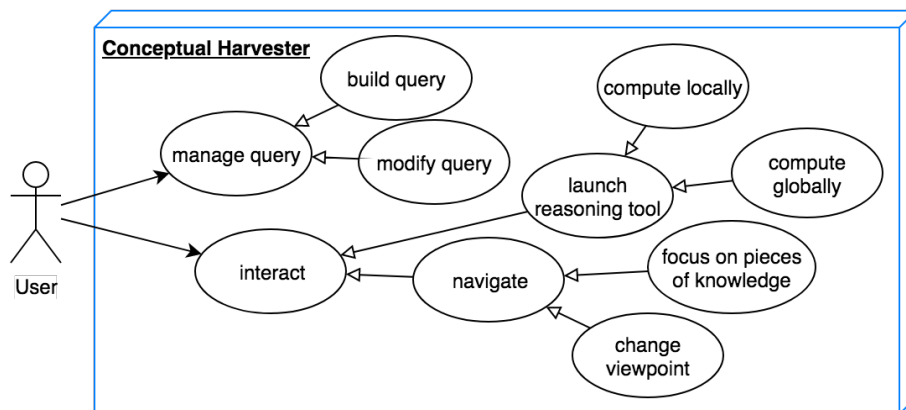


Figure 4. Diagramme de cas d'utilisation du système *Conceptual Harvester* en utilisant le langage UML.

En préalable à l'analyse des connaissances de Knomana, l'utilisateur effectue une navigation décrite au moyen du diagramme de séquences présenté à la FIGURE 5. D'après celui-ci, l'utilisateur réduit tout d'abord la liste des connaissances aux espèces du genre *Fusarium*, i.e. à 457 connaissances (*extract (Fusarium)* sur la figure). L'utilisateur exécute alors RCAExplore (*compute lattice*) qui calcule quatre treillis inter-connectés. Les treillis organisant les objets de type plante de protection, organisme cible, système protégé et connaissance comportent respectivement 20, 7, 5 et 55 concepts. Puis, pour accéder exclusivement aux connaissances sur le *Fusarium oxysporum*, il réduit la zone de recherche à cette espèce (*focus (fusarium oxy)*), puis au Burkina Faso (*focus (burkina)*). En l'absence de réponse à Q1 dans les deux cas, il exécute enfin l'algorithme de raisonnement (*recherche(...)*) dans la perspective d'obtenir une solution.

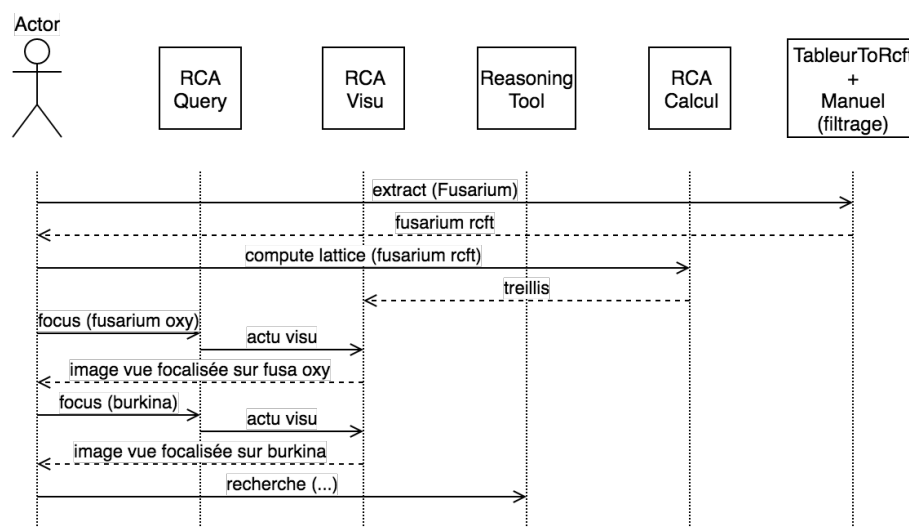


Figure 5. Diagramme de séquences utilisant les composants actuels pour le scénario d'utilisation suivant : « Quelle plante pesticide pourrait combattre le *Fusarium oxysporum* au Burkina Faso ? ».

L'objectif de ce nouvel algorithme de raisonnement (algorithme 1) est d'extraire une réponse du treillis en examinant les points communs cachés entre deux concepts. La FIGURE 6 illustre la problématique de recherche dans le treillis pour Q1. Cette figure montre que les connaissances sur le *Fusarium oxysporum* (concept 28) ont toutes pour attributs communs :

- la plante protégée est l'*ArachisHypogaea (ArHy)* sous forme de denrées stockées (*onStock*).
- la partie de la plante pesticide utilisée est la feuille (*useFeuille*).
- la plante pesticide est toxique pour le *Fusarium oxysporum (interractToxique)*.
- les connaissances proviennent du Bénin (*paysBénin*).

Ce concept 28 hérite des concepts 29 (pays = Bénin) et 30 (interaction = toxicité). Le Burkina Faso (concept 20) se trouve sur une branche du treillis qui n'a pour concepts communs avec le concept 28 que la borne supérieure et la borne inférieure du treillis. Il s'ensuit que les concepts 20 et 28, introduisant respectivement Burkina Faso et *Fusarium oxysporum*, n'ont aucun lien commun.

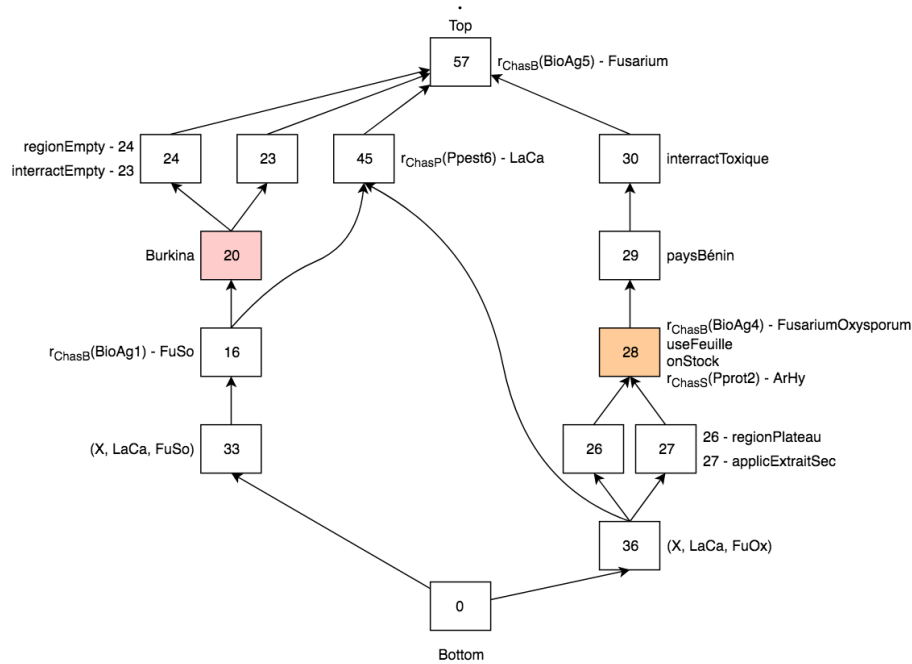


Figure 6. Treillis des connaissances simplifié contenant les éléments pertinents pour le scénario « *Fusarium oxysporum* au Burkina Faso ». La flèche incurvée indique le raisonnement à produire pour accéder à la solution.

L'accès à la solution de Q1 (concept 45) nécessite un raisonnement particulier. L'algorithme 1 recherche dans le treillis les points communs cachés entre deux concepts. Pour des questions de facilité de compréhension, les notations des contextes \mathcal{K}_P , \mathcal{K}_B et \mathcal{K}_S sont adaptées à la question Q1.

Avant la description de l'algorithme, voici les définitions et notations des différentes notions utilisées. Un contexte formel \mathcal{K} est un 3-tuple (G, M, I) , où G est un ensemble d'objets, M un ensemble d'attributs et $I \subseteq G \times M$ une relation objet-attribut. Par exemple, sur la FIGURE 1 qui représente la relation ternaire de l'ontologie, le contexte \mathcal{K}_B décrit les objets bioagresseurs par leur taxonomie (espèce, genre, famille). La FIGURE 1, sur laquelle on peut voir 4 contextes et 3 relations, présente une famille de contextes relationnels qui est une paire (\mathbf{K}, \mathbf{R}) , où \mathbf{K} est un ensemble de contextes formels et \mathbf{R} un ensemble de relations r entre les contextes formels. La source (ou le domaine) de la relation est notée $\text{dom}(r)$, et la destination (ou la portée, ou encore la cible) est notée $\text{ran}(r)$. Un concept est formé

à partir d'un contexte formel, ce dernier pouvant en former plusieurs. Un concept $C = (Extent(C), Intent(C))$ regroupe, dans son extension, un ensemble maximal d'objets partageant un ensemble maximal d'attributs (son intension). Pour l'écrire formellement, $Extent(C) = \{g \in G \mid \forall m \in Intent(C), (g, m) \in I\}$ et $Intent(C) = \{m \in M \mid \forall g \in Extent(C), (g, m) \in I\}$.

Le treillis formé à partir d'un contexte formel est noté $\mathcal{L}_{\mathcal{K}}$. Il organise les concepts selon un ordre partiel de spécialisation \preceq_C basé sur l'inclusion des extensions (en remontant) ou des intensions (en descendant). Pour deux concepts C_1 et C_2 , $C_1 \preceq_C C_2$ est représenté visuellement par le fait que C_1 soit au-dessous de C_2 dans la représentation graphique. Un concept est dit introducteur d'un attribut a (resp. d'un objet o), noté $\mu(a)$ (resp. $\gamma(o)$) s'il est, en partant du haut (resp. du bas), le premier concept à posséder cet attribut (resp. cet objet). Un attribut relationnel est une expression $qr(C)$, où q est un quantifieur, r est une relation et C est un concept formel cible de la relation r , c'est-à-dire que l'extension de C contient des objets de $ran(r)$. Une relation r , appliquée sur son contexte formel source $\mathcal{K} = (G, M, I)$ avec un quantifieur q , produit un nouveau contexte « étendu » appelé extension relationnelle partielle (G^+, M^+, I^+) , où $G^+ = G$, M^+ contient les attributs relationnels visant le contexte formel cible et I^+ contient les relations entre les objets et les attributs relationnels.

L'algorithme admet en entrée la famille de contextes relationnels, les treillis et les autres paramètres suivants :

- un attribut m du contexte formel principal $\mathcal{K}_C = (G_{\mathcal{K}_C}, M_{\mathcal{K}_C}, I_{\mathcal{K}_C})$, le *Burkina Faso*,
- un attribut relationnel a_{relg} vers le concept introducteur de l'objet g dans \mathcal{K}_B , g étant le *Fusarium oxysporum*, par la relation r_{ChasP} ,
- la relation r ayant comme source le contexte principal et comme cible le contexte de l'ensemble d'objets que l'on cherche, ici r_{ChasP} .

Les lignes 2 et 3 de l'algorithme récupèrent les concepts introducteurs $\mu m = \mu(m)$ et $\mu a_{relg} = \mu(a_{relg})$ dans \mathcal{K}_C , en l'occurrence les concepts 20 et 28 respectivement. Aux lignes 4 et 5, Em et Eg récupèrent la liste des objets des extensions de μm et μa_{relg} . Les itérations imbriquées des lignes 6 à 12 récupèrent dans \mathcal{A}_{Rel} les attributs relationnels de M_C^+ concernés par la relation r d'intérêt, c'est-à-dire vers des concepts de la cible de la relation, et possédés en même temps par un objet de Em et un objet de Eg en comparant les listes d'attributs de chaque objet deux à deux. À la ligne 13, \mathcal{C}_{Rel} récupère les concepts cibles des attributs de \mathcal{A}_{Rel} . Pour Q1, on récupère donc $\mathcal{C}_{Rel} = \{Concept_Ppest_16, Concept_Ppest_6, Concept_Ppest_17, Concept_Ppest_14, Concept_Ppest_15\}$. Parmi eux, on ne garde dans $\mathcal{C}_{RelIntro}$ que les concepts introducteurs d'objets grâce à l'itération des lignes 14 à 16. Puis l'ensemble \mathcal{X} des concepts les plus petits de $\mathcal{C}_{RelIntro}$ est retourné selon l'ordre de spécialisation \preceq_C . Dans notre exemple, $\mathcal{X} = \{Concept_Ppest_6\}$. Enfin, les objets des concepts de \mathcal{X} , c'est-à-dire les éléments de $\{obj \mid \exists C \in \mathcal{X} \text{ et } obj \in Extent(C)\}$, sont mis en évidence visuellement et présentés à l'utilisateur. Vis-à-vis de la question Q1, la solution proposée par l'algorithme est *Lantana camara* (LaCa) car cette

Algorithme 1 : HiddenPoint

Data : famille de contextes relationnels, $m \in M_{\mathcal{K}_C}, a_{relg} \in M^+_{\mathcal{K}_C}$ avec
 $g \in G_{\mathcal{K}_B}, r \subseteq G_C \times G_P, \mathcal{L}_{\mathcal{K}_C}, \mathcal{L}_{\mathcal{K}_B}, \mathcal{L}_{\mathcal{K}_P}$

Result : $\mathcal{X} = \{x \mid x \in G_{\mathcal{K}_P}\}$ un ensemble d'objets

- 1 $\mathcal{A}_{Rel} := \emptyset; \mathcal{C}_{RelIntro} := \emptyset;$
- 2 $\mu m := \text{calculConceptIntroAtt}(\mathcal{L}_{\mathcal{K}_C}, m);$
- 3 $\mu a_{relg} := \text{calculConceptIntroAtt}(\mathcal{L}_{\mathcal{K}_C}, a_{relg});$
- 4 $E_m := \text{Extent}(\mu m);$
- 5 $E_g := \text{Extent}(\mu a_{relg});$
- 6 **foreach** obj_m of E_m **do**
 - // recup la liste des attributs relationnels de obj_m
 - 7 $A' := \{a \in M_C^+ \mid (obj_m, a) \in I_C^+\};$
 - 8 **foreach** obj_g of E_g **do**
 - // recup la liste des attributs relationnels de obj_g
 - 9 $B' := \{a \in M_C^+ \mid (obj_g, a) \in I_C^+\};$
 - 10 **foreach** $attrib$ of A' **do**
 - 11 **if** $(attrib \in B') \ \& \ (attrib \in \text{ran}(r))$ **then**
 - 12 $\mathcal{A}_{Rel} := \mathcal{A}_{Rel} \cup \{attrib\};$
 - // recup concepts cibles des attributs relationnels
 - 13 $\mathcal{C}_{Rel} := \{C \mid \exists a = (\exists r(C)) \text{ avec } a \in \mathcal{A}_{Rel}\};$
 - 14 **foreach** $concept$ of \mathcal{C}_{Rel} **do**
 - 15 **if** $\text{estConceptIntroObj}(concept)$ **then**
 - // $concept \in \mathcal{C}_{Rel} \mid \exists obj \in G_{\mathcal{K}_F}$ avec $concept = \gamma(obj)$
 - 16 $\mathcal{C}_{RelIntro} := \mathcal{C}_{RelIntro} \cup \{concept\};$
 - 17 $\mathcal{X} := \min_{\leq_C}(\mathcal{C}_{RelIntro});$
 - 18 **return** $\mathcal{X};$

plante est utilisée au Bénin pour traiter le *Fusarium oxysporum* (connaissance (X, LaCa, FuOx) du concept 33) et au Burkina Faso pour lutter contre le *Fusarium solani* (connaissance (X, LaCa, FuSo) du concept 36), une autre espèce de champignon du genre *Fusarium*. Cette solution met en exergue l'usage de la classification du vivant dans le raisonnement.

4. Problématique de recherche et travaux connexes

La problématique de recherche de ce travail est de naviguer et d'explorer une base de connaissances au moyen de l'ACF, notre cas d'application étant la base Knomana. Plusieurs sous-problématiques ont été identifiées en regard de la littérature.

Des travaux antérieurs dans le domaine de la santé et de l'environnement ont déjà mobilisé l'ACF, à l'exemple de (Jay *et al.*, 2006) pour concevoir des réseaux de santé efficaces et planifier les ressources médicales à partir de l'analyse de flux de patients, (Keller *et al.*, 2012) pour évaluer la complexité des relations entre des maladies ré-

nales, et (Ozcan-Deniz, Zhu, 2014) pour concevoir des édifices dans un objectif socio-écologique pérenne. D'autres ont mobilisé l'ACR, à l'exemple de (Nica *et al.*, 2018) et (Ouzerdine *et al.*, 2019) pour évaluer des cours d'eau dans un objectif de restauration de la biodiversité. La première sous-problématique de notre travail sera d'évaluer l'aptitude de l'ACF et de ses extensions relationnelles à classer autant d'informations.

L'ACR peut être utilisée pour différents types de raisonnement. L'exemple du cas d'utilisation présenté en section 3 vise à identifier au sein du treillis un objet au moyen d'un raisonnement particulier. D'autres méthodes de classification sont présentées dans la littérature, à l'exemple des arbres de décision et des règles d'association (Han *et al.*, 2011). Ces méthodes possèdent des propriétés de classification différentes de celles de l'ACF (Ouzerdine, 2018). Dans le cadre de l'ACF, l'approche Graph-FCA (Ferré, 2015) traite de données relationnelles en considérant des graphes de connaissances basés sur des relations n-aires. Dans ce cadre, l'intension d'un concept est implémentée sous forme de *graph pattern* et l'extension d'un ensemble de tuples. La seconde sous-problématique de ce travail sera par conséquent d'identifier la méthode de classification qui satisfasse les objectifs de classification et de visualisation de l'utilisateur.

La structure logicielle adoptée pour ce travail repose sur une architecture de composants interconnectés. Celle-ci offre l'avantage de pouvoir échanger le contenu des composants sans altérer le fonctionnement du système. Dans le cas du composant Reasoning Tool du moissonneur conceptuel, l'intégration du logiciel d'analyse Weka⁶, par exemple, permettrait d'accéder aux différentes méthodes de classification introduites ci-dessus. Le composant Knowledge Base permettra également de naviguer/explore des bases de connaissances de format différent, e.g. triple store ou base de données de type graphe. Quel que soit le modèle de structuration des connaissances adopté, celui-ci sera décrit au moyen d'une ontologie. Les ontologies peuplées présentes dans AgroPortal⁷ (Jonquet *et al.*, 2018) peuvent permettre d'ailleurs d'enrichir les connaissances de la base Knomana. Dans le cas de Knomana, la structuration des connaissances comporte une relation ternaire. Cette structure a été transformée en un ensemble de relations binaires, par un expert du domaine, pour permettre la classification des données par l'ACF. La troisième sous-problématique que pose ce travail sera de disposer d'un système permettant de transformer toute structure d'ontologie en un ensemble de contextes formels.

Enfin, le moissonneur conceptuel vise à être utilisable par des utilisateurs différents pour répondre à des questions (préoccupations) diverses. Ce faisant, l'utilisateur occupe une place prépondérante dans ce système. Aussi, que ce soit pour la formulation de la requête ou pour la visualisation des résultats, le choix des idiomes est important. Concernant les prototypes logiciels utilisant l'ACF pour la classification, les résultats sont visuellement présentés sous forme de graphe, ou encore de nuage de mots (Keip, 2018). Dans un autre cadre, (Cuenca, 2018) a représenté des données comportant dif-

6. <https://www.cs.waikato.ac.nz/ml/weka/>

7. <http://agroportal.lirmm.fr/>

férents niveaux d'organisation hiérarchique au moyen de plusieurs idiomes interactifs et interconnectés. La formulation de la requête s'effectue avec un éditeur de graphe. Ces types d'idiomes sont pratiques pour un initié, mais s'appliquent beaucoup moins aux producteurs agricoles car ils ne sont pas intuitifs. La quatrième et dernière sous-problématique de ce travail consistera donc à identifier des idiomes en fonction des cas d'utilisation et des utilisateurs cibles.

5. Conclusion

Le travail présenté dans cet article concerne la navigation et l'exploration de la base de connaissances Knomana, afin de répondre à des questions de santé animale et végétale. L'implémentation des méthodes de navigation et d'exploration se fera dans une structure logicielle reposant sur une architecture constituée du composant Knowledge Base (comportant la base Knomana) relié au composant Conceptual Harvester permettant sa navigation/exploration. Ce dernier comporte à son tour (i) le composant Query builder pour formuler les requêtes, (ii) le Knowledge extractor pour récupérer les connaissances extraites du Query builder d'après la requête, (iii) le Reasoning Tool pour classer les connaissances et (iv) le Knowledge viewer pour permettre à l'utilisateur de naviguer/explorer la classification. Le Reasoning Tool comportera des outils de raisonnement tels que RCAExplore et Weka, mais également des algorithmes de raisonnement pour répondre à des cas d'utilisation particuliers tels que celui présenté dans cet article.

Notre problématique consiste à déterminer une méthode pour naviguer et explorer une base de connaissances au moyen de l'ACF. Pour la résoudre, il faudra au préalable répondre à quatre sous-problématiques. La première concerne l'évaluation de l'aptitude de l'ACF et de ses extensions ACR et Graph-FCA à classer un grand volume de données, à savoir 30 000 connaissances décrites au moyen de 36 informations. La seconde portera sur l'identification des méthodes de classification pertinentes et au développement des algorithmes de raisonnement nécessaires pour différents cas d'utilisation. La troisième concerne le lien entre le modèle de l'ontologie utilisée pour structurer les connaissances en regard de la forme des données d'entrée attendue par les algorithmes de classification. Enfin, la dernière portera sur l'identification des idiomes d'après les cas d'utilisation et les acteurs cibles.

Remerciements

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme d'Investissements d'Avenir portant la référence ANR-16-CONV-0004.

Bibliographie

Bazin A., Carbonnel J., Huchard M., Kahn G. (2018). On-demand relational concept analysis. *CoRR*, vol. abs/1803.07847. Consulté sur <http://arxiv.org/abs/1803.07847>

- Cuenca E. (2018). *Visualisation de données dynamiques et complexes : des séries temporelles hiérarchiques aux graphes multicouches*. Thèse de doctorat, Université de Montpellier. (<http://www.theses.fr/s174769>)
- Dolques X., Ber F. L., Huchard M., Nebut C. (2013). Analyse relationnelle de concepts pour l'exploration de données relationnelles. In C. Vrain, A. Péninou, F. Sèdes (Eds.), *Actes de la conférence Extraction et Gestion des Connaissances (EGC'2013), 29 janvier - 01 février 2013, Toulouse, France*, vol. RNTI-E-24, p. 121–132. Hermann-Éditions. Consulté sur <http://editions-rnti.fr/?inprocid=1001829>
- Ferré S. (2015). A proposal for extending formal concept analysis to knowledge graphs. In *Proc. of the 13th International Conference on Formal Concept Analysis (ICFCA'15), Nerja, Spain, June 23-26*, vol. 9113, p. 271–286. Springer.
- Ganter B., Wille R. (1999). *Formal concept analysis - mathematical foundations*. Springer.
- Hacene M. R., Huchard M., Napoli A., Valtchev P. (2013). Relational concept analysis: mining concept lattices from multi-relational data. *Ann. Math. Artif. Intell.*, vol. 67, n° 1, p. 81–108. Consulté sur <https://doi.org/10.1007/s10472-012-9329-3>
- Han J., Kamber M., Pei J. (2011). Data mining concepts and techniques. In, 3rd éd., vol. 673, chap. 8,9. Morgan Kaufmann.
- Jay N., Kohler F., Napoli A. (2006). Using Formal Concept Analysis for Mining and Interpreting Patient Flows within a Healthcare Network. In *Proc. of the 4th Int. Conference on Concept Lattices and Their Applications, (CLA'06), Tunis, Tunisia, October 30 - November 1, 2006, Selected Papers*, p. 263–268.
- Jonquet C., Toulet A., Arnaud E., Aubin S., Yeumo E. D., Emonet V. *et al.* (2018). Agroportal: A vocabulary and ontology repository for agronomy. *Computers and Electronics in Agriculture*, vol. 144, p. 126 - 143. Consulté sur <http://www.sciencedirect.com/science/article/pii/S0168169916309541>
- Keip P. (2018). *Navigation conceptuelle et exploration pour la sélection de plantes pesticides et antibiotiques en santé animale et végétale dans les pays du sud*. Mémoire de Master, Université de Montpellier.
- Keller B. J., Eichinger F., Kretzler M. (2012). Formal concept analysis of disease similarity. *AMIA Summits on Translational Science Proceedings*, vol. 2012, p. 42.
- Munzner T. (2014). *Visualization analysis and design*. A K Peters. Consulté sur <http://www.cs.ubc.ca/~tmm/vadbook/>
- Nica C., Braud A., Ber F. L. (2018). Exploring heterogeneous sequential data on river networks with relational concept analysis. In P. Chapman, D. Endres, N. Pernelle (Eds.), *Proc. of the 23rd International Conference on Conceptual Structures, Graph-Based Representation and Reasoning (ICCS'18), Edinburgh, UK, June 20-22*, vol. 10872, p. 152–166. Springer. Consulté sur https://doi.org/10.1007/978-3-319-91379-7_12
- Ouzerdine A. (2018). *Étude de méthodes d'analyse de données et d'extraction de connaissances pour l'usage des plantes pesticides, Projet KNOMANA*. Mémoire de Master, Université de Montpellier.
- Ouzerdine A., Braud A., Dolques X., Huchard M., Ber F. L. (2019). Régler le processus d'exploration dans l'analyse relationnelle de concepts - le cas de données hydroécologiques. In M. Rousset, L. Boudjeloud-Assala (Eds.), *Actes de la conférence Extraction et Gestion*

des Connaissances, (EGC'19), Metz, France, January 21-25, 2019, vol. E-35, p. 57–68. Hermann-Éditions. Consulté sur <http://editions-rnti.fr/?inprocid=1002468>

Ozcan-Deniz G., Zhu Y. (2014). Application of formal concept analysis in estimating the cost of green buildings. In *Proc. of the 50th International Conference of the Associated Schools of Construction (ASC'14)*.

Zeleny M. (2013). Integrated knowledge management. *International Journal of Information Systems and Social Change (IJISSC)*, vol. 4, n° 4, p. 62–78.